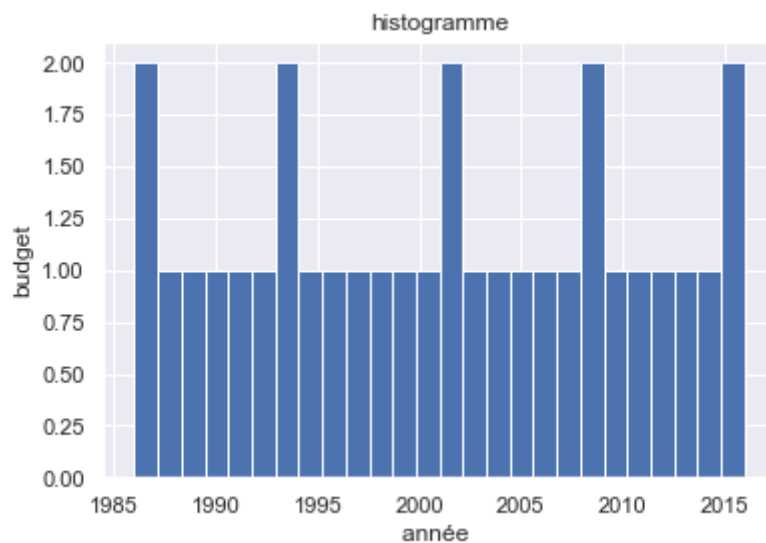


## Problématique :

Depuis une analyse sur 26 ans nous voudrions connaître la moyenne des budgets investis dans les films afin de pouvoir établir une prévision sur les budgets alloués l'an prochain.

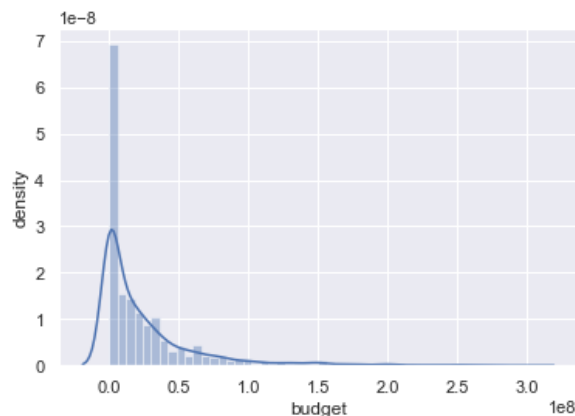
## Partie 1 : Analyse graphique des données

Tout d'abord nous avons commencé par faire un histogramme de la somme des budgets par année. Ci-dessous capture de l'histogramme :



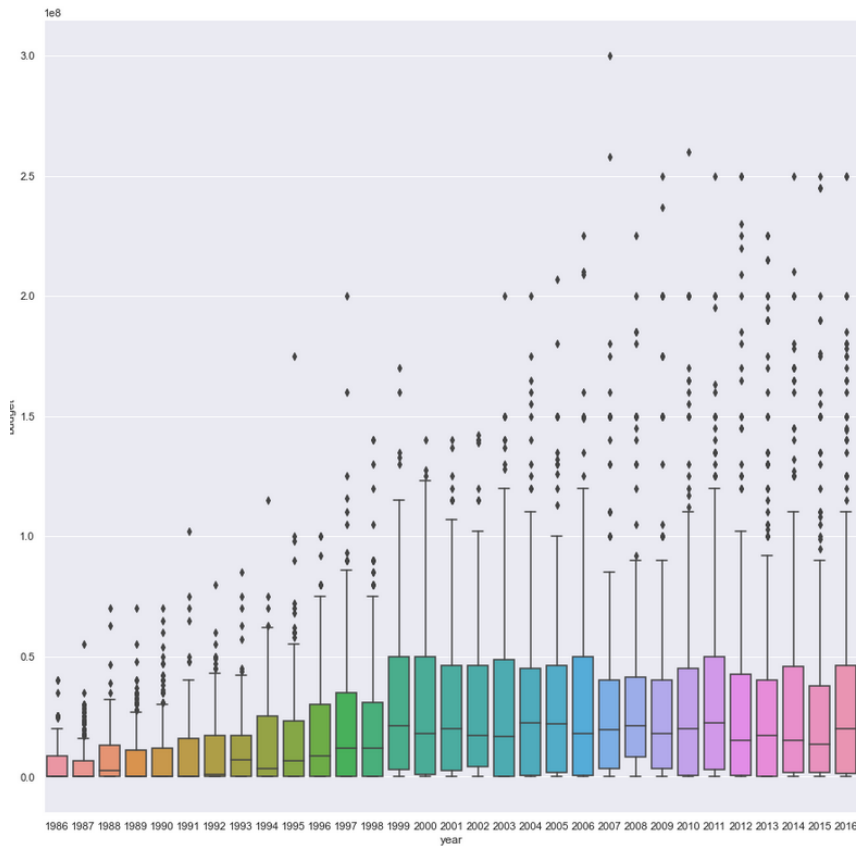
### A. Diagrammes de répartition des données

Le diagramme suivant est appelé le diagramme en cloche ou diagramme de gauss celui-ci nous permet de d'évaluer la densité moyenne du budget alloué par année.



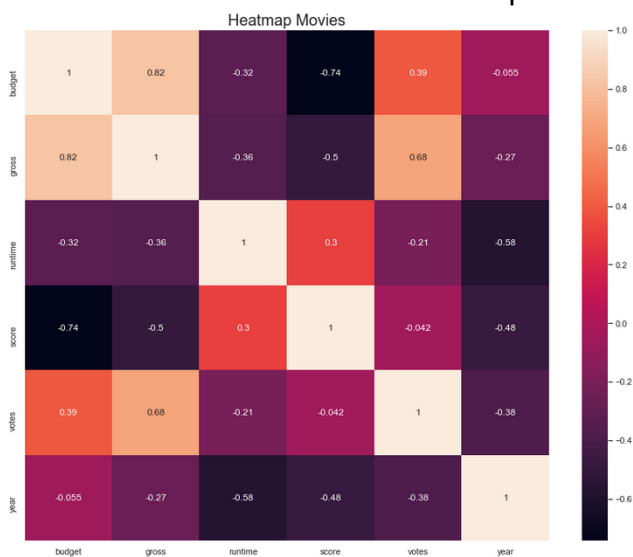
## B. Boîtes à moustache

Le diagramme à moustache nous permet de représenter quelques indicateurs de position du caractère étudié à savoir la médiane, l'équartiles, le minimum et le maximum



## C. Heatmap + observations sur les corrélations

La Heatmap est une représentation graphique de la donnée ou les valeurs contenues dans une matrice sont représenté sous différentes couleurs



## Partie 2 : Model Building

### A. 2 algorithmes avec 2 paramètres différents

La prise de décision a été calculé avec un arbre de décision.

**Définition arbre de décision :**

Arbre de décision : est de créer un modèle de prédiction de valeurs d'une variable ciblée en utilisant une règle de décision en fonction des données passées en entrée

Dans notre arbre de décision nous obtenons un seuil de 50.7%

Si l'hypothèse est située dans les 50.7% l'hypothèse sera acceptée si celle-ci dépasse le seuil elle sera donc rejetée

La confirmation de notre variable indépendante a été calculée avec une régression linéaire.

**Définition arbre de décision :**

Régression linéaire la régression linéaire est utilisée afin d'étudier les liens entre une variable dépendante et des variables indépendantes afin de juger de la qualité de la variable dépendante par les variables indépendantes Dans le cas d'un mini projet nous avons voulu établir une prédiction quant au budget qu'il sera dépensé sur les films

Notre régression linéaire nous a sorti un score de 50,9%

Dans notre cas notre variable indépendante qui est le budget alloué sera confirmé à 50,9% par notre algorithme.

### Affichage des coefficients/ accuracy

Accuracy\_score :

La précision exprime le coefficient d'exactitude des prédictions.

Le coefficient de régression indique de par sa valeur le taux de croissance de la fonction :

- Si le coefficient est positif cela indique une croissance
- Si le coefficient est négatif cela indique une décroissance

Dans notre étude nous avons un taux d'exactitude de 50.9% notre coefficient est positif ce qui indique donc une croissance.

## Partie 3 : Features Importance

### A. Affichage sous forme de Histogramme

Les Features importance nous permet de mesurer l'impact de chaque variable lors de la prédiction.

Dans l'histogramme ci-dessous nous pouvons observer que la Feature 0 (L'année) a un impact de 38% et la Features 1 (le box-office) à un impact de 62% sur le budget.

```
Feature: 0, Score: 0.38  
Feature: 1, Score: 0.62
```

